Les questionnaires dans la démarche de prévention du stress au travail

Le questionnaire fait partie des outils d'évaluation qu'il est possible d'utiliser dans une démarche de prévention du stress au travail. Il permet à l'entreprise de mesurer l'ampleur du phénomène au niveau de l'ensemble des salariés, grâce à un recueil d'informations précises sur les sentiments, les perceptions et les expériences de ces derniers, pour un coût relativement modéré et de manière assez rapide. Il existe de nombreux questionnaires qui ne mesurent pas tous les mêmes éléments. Cet article a pour but de donner quelques pistes au préventeur pour l'aider à choisir celui ou ceux pertinent(s) au regard de la situation à évaluer.

V. LANGEVIN*, M. FRANÇOIS**, S. BOINI***, A. RIOU*

** Département Expertise et conseil technique, INRS*
*** Département Homme au travail, INRS*
**** Département Épidémiologie en entreprise, INRS*

En résumé

Les questionnaires sont souvent utilisés dans les démarches de diagnostic et de prévention du stress et des risques psychosociaux au travail (ou RPS). Leurs construction et utilisation nécessitent de respecter un certain nombre de règles détaillées dans cet article, notamment en ce qui concerne les qualités psychométriques de validité, fidélité et sensibilité.

La place des questionnaires dans la démarche de prévention du stress est rappelée.

Une analyse comparative des différents outils d'évaluation disponibles en français, parmi ceux les plus utilisés dans les études sur le stress et les risques psychosociaux au travail, a été menée. Ces outils sont décrits au travers d'un certain nombre de critères. Les analyses de 36 questionnaires seront publiées sous forme de fiches portant la référence FRPS dans la revue *Documents pour le Médecin du Travail.*

Pour le salarié, répondre à un questionnaire anonyme et confidentiel c'est avoir la possibilité de s'exprimer sur ses conditions de travail. Son avis et sa perception sont indispensables pour approcher la réalité de son poste de travail, il en estime ainsi les contraintes, les dysfonctionnements, les nuisances physiques, les sources de satisfaction…

Toutefois, compte tenu du fait que les risques psychosociaux, et plus particulièrement le stress, sont très présents dans le débat social, politique et médiatique, il est nécessaire de compléter l'utilisation de questionnaires par d'autres sources d'évaluation qui permettent de limiter l'influence du contexte social sur la tonalité des réponses. La confrontation des informations obtenues par questionnaire avec celles provenant d'autres sources (observation du travail, analyse des données administratives, prise en compte des données médicales, indicateurs physiologiques, entretiens collectifs, etc.) validera (ou non) la qualité de ces informations et favorisera l'émergence de solutions adaptées aux problèmes soulevés.

Il n'en reste pas moins que le recours aux questionnaires est particulièrement indiqué pour des entreprises qui emploient au minimum une centaine de personnes ou pour des petites entreprises réalisant la même activité et pouvant être rassemblées au sein d'une seule enquête. Si une entreprise emploie une trentaine de salariés (par exemple), d'autres méthodes de recueil de données sont plus indiquées (entretiens individuels et/ou collectifs, observations du travail…).

Définitions et aspects méthodologiques

DÉFINITIONS

Le questionnaire est un outil de recueil d'informations, basé sur la formulation de questions ou d'affirmations, couramment utilisé dans les méthodes d'investigation en sciences humaines, sociales et épidémiologiques (au même titre que les entretiens notamment) ou encore dans le marketing ou la presse. De ce



fait, le choix du questionnaire dépend du contexte et des objectifs poursuivis par les enquêteurs.

L'enregistrement des réponses est la plupart du temps prédéterminé (questions fermées). Il existe également des questions dites « ouvertes », généralement en fin d'un questionnaire « fermé ». Ces questions ouvertes viennent explorer un ou plusieurs aspects de la réalité insuffisamment ou non pris en compte par les questionnaires retenus pour l'étude. Le traitement des réponses à ces questions est différent de celui des réponses aux questions fermées.

On distingue les questionnaires auto-administrés et ceux de pré-diagnostic, dits « d'hétéro-évaluation », généralement utilisés en groupe de travail, renseignés de façon collégiale et utilisés en amont d'un véritable diagnostic sur les situations de travail.

Quand un questionnaire est construit à partir d'un modèle théorique, selon une méthodologie rigoureuse, et qu'il aboutit à un score, il est appelé « échelle de mesure ». Un questionnaire mesurant plusieurs dimensions et comportant plusieurs échelles est parfois appelé « inventaire » ou « échelle multidimensionnelle ». La plupart du temps, les modalités de réponses sont graduées, exprimant une fréquence, une intensité, un degré d'accord (modalités de réponse sur une échelle dite de « Likert »). Parfois, les réponses sont simplement dichotomiques en « oui » ou « non ».

Le principe de construction des échelles de mesure permet d'appliquer des techniques statistiques particulières visant à vérifier les qualités psychométriques de ces instruments de mesure que sont la validité, la fidélité (ou fiabilité) et la sensibilité *(annexe 1)*.

Les échelles de mesure permettent de recueillir des informations sur des opinions, des attitudes, des sentiments, des perceptions de l'environnement ou de soi-même. En psychologie, elles sont également utilisées pour mesurer des aptitudes cognitives.

Ce type de mesure s'appuie sur une théorie de la mesure en psychologie (« théorie du score vrai ») dont les principes essentiels sont les suivants :

- un individu possède des traits (des caractéristiques personnelles) relativement stables ;

- le score obtenu (imparfait) X est la résultante d'un score « vrai » T et d'une part d'erreur *e*. Ainsi, $X = T + e$;

- lors de la répétition de la mesure sur un même individu, les erreurs se répartissent autour du score « vrai » et leur moyenne est nulle.

Pour connaître une caractéristique d'une personne, que ce soit son poids, une représentation mentale ou une caractéristique psychologique, plus il y a de mesures, plus la moyenne de celles-ci fournit une approximation satisfaisante de cette caractéristique, si les réponses sont indépendantes. L'existence, dans les échelles de mesure, de plusieurs items explorant une même facette est une façon de s'approcher du score vrai (et donc limiter l'erreur de mesure).

Une grande partie des questionnaires utilisés pour explorer les problèmes de stress sont des échelles de mesure, composées de plusieurs questions ou affirmations (aussi appelées items) avec plusieurs modalités de réponses proposées.

CALCUL DE SCORE

Les réponses apportées donnent lieu au calcul d'un score selon des modalités variables dont la plus simple et la plus répandue consiste à effectuer la somme des points attribués aux différentes réponses d'une même échelle. Le calcul peut également être plus sophistiqué en introduisant une pondération des items, décidée *a priori* ou *a posteriori* par les auteurs pour donner le même poids aux différents items d'une échelle ou entre des échelles différentes d'un même inventaire (exemples : *Job Content Questionnaire*, plus couramment appelé questionnaire de Karasek, ou le *Nottingham Health Profil*).

Les échelles de mesure peuvent se subdiviser en sous-échelles, comme dans le questionnaire de Karasek où l'échelle « latitude décisionnelle » se subdivise en deux sous-échelles, « développement des compétences » et « autonomie de décision », ce qui permet de calculer un score et deux sous-scores.

ÉTALONNAGE ET INTERPRÉTATION DES SCORES

Une fois les scores calculés, il s'agit de les interpréter.

Pour faciliter l'interprétation et la comparaison des scores, il arrive que les scores bruts soient transformés en scores (ou notes) standardisé(e)s. Cette standardisation s'effectue la plupart du temps selon le modèle gaussien (loi normale) ou selon les percentiles.

Par exemple, un score de 13 à l'échelle « exigences psychologiques » du questionnaire de Karasek ne livre aucune information en soi. Pour lui donner du sens, il est nécessaire de savoir si c'est un score plutôt moyen, faible ou élevé. Pour cela, il faut comparer ce score de 13 à celui obtenu dans un échantillon représentatif de la population générale, appelé échantillon normatif ou étalonnage, servant à produire des valeurs de référence. Dans le cas présent, le score de 13 peut être comparé aux données obtenues dans le cadre de l'enquête SUMER, qui constitue un large échantillon de la population active française. Dans cet échantillon normatif, le score moyen observé sur l'échelle « exigences psychologiques » est 21,86 et l'écart-type est de 4,40. Le score de 13 se situe donc à peu près à deux écarts-types en

dessous de la moyenne, ce qui dans une distribution normale (modèle gaussien) est un score très faible (seulement 2,5 % de l'échantillon de référence ont obtenu un score plus faible que 13).

Disposer d'un étalonnage de référence est donc fortement utile pour interpréter les scores. Selon Kline, un échantillon valide est un échantillon représentatif et de taille suffisante pour éviter les erreurs standard de statistiques descriptives. Par définition, le choix des sujets composant cet échantillon pose la question de la généralisation des résultats, ceux-ci devant être pertinents pour la population cible, sans quoi la présence de valeurs de référence ne serait d'aucune utilité et pourrait conduire à une mauvaise interprétation des résultats.

En l'absence de cet étalonnage, les scores obtenus en entreprise pourront être interprétés par comparaison entre les différents services, ateliers ou secteurs de cette entreprise.

De manière complémentaire, une autre approche consiste à faire appel à des critères externes ayant du sens pour la personne interrogée ou le médecin (*anchor-based*), et à essayer de définir des différences cliniquement significatives.

Plusieurs types d'information peuvent aider à l'interprétation des scores :

- données comparatives sur la distribution de scores obtenus à partir des diverses populations, incluant si possible un échantillon représentatif de la population cible ;
- résultats de nombreuses études utilisant l'outil en question, permettant ainsi une familiarisation avec celui-ci et en facilitant l'interprétation ;
- relation entre les scores (ou les changements de scores) et des conditions reconnues comme cliniquement importantes ;
- relation entre les scores (ou les changements de scores) et des conditions reconnues comme socialement importantes ;
- relation entre les scores (ou les changements de scores) et la notion de différence minimale perçue comme significative par les personnes enquêtées ;
- valeur pronostique des scores pour des événements reconnus comme importants.

Recensement et analyse des outils

Un recensement des divers questionnaires et échelles de mesure utilisés dans les études sur le stress au travail a été réalisé à partir d'une revue de la littérature internationale sur le stress et les risques psychosociaux et de la littérature grise disponible à l'INRS, entre 2006 et 2009.

Sans prétendre à l'exhaustivité, il a permis d'identifier 55 outils, de langue française et anglaise. Seuls ont été retenus pour analyse ceux ayant une version en français, soit 36 outils (donnant lieu à autant de fiches signalétiques).

Plusieurs critères ont été retenus pour les décrire et les classer en 9 catégories, selon leurs objectifs *(tableau I)* :

- Repérage par des tiers de situations de travail stressantes : une personne extérieure à la situation de travail en fait l'évaluation, par exemple l'encadrement de proximité, le médecin du travail. Ce sont des outils d'évaluation de pré-diagnostic en hétéro-évaluation ;
- Situation de travail perçue : la personne travaillant au poste évalue ses propres conditions de travail ;
- État de stress perçu : identification des mécanismes psychocognitifs et du sentiment de contrôle vécu dans des situations stressantes, dans la vie quotidienne ou pendant le travail ;
- Stratégies d'adaptation au stress : repérage des stratégies d'adaptation les plus fréquemment utilisées par les personnes ;
- Symptômes de stress : identification et évaluation des différents symptômes de stress ressentis, par auto – évaluation ;
- Atteinte du rapport psychologique au travail : satisfaction, stimulation… ;
- Atteinte à la santé physique et mentale ;
- Violences internes au travail ;
- Questionnaires transversaux comprenant un certain nombre d'échelles portant sur trois ou plus des points précédents.

Chaque fiche signalétique comprend 18 rubriques, chacune renseignée par l'analyse d'articles et de documents scientifiques se rapportant à l'outil étudié et référencés dans la dernière rubrique *(tableau II)*.

L'ensemble des 36 fiches sera publié dans la revue les *Documents pour le Médecin du Travail* sur plusieurs numéros. L'ordre de parution a été déterminé d'une part sur la notoriété des outils, d'autre part sur la nécessité de ne pas dissocier des outils allant ensemble et enfin le souhait de mettre à disposition des outils d'objectifs différents dans chaque numéro [(1)].

(1) une première série d'analyses est publiée dans ce même numéro pp. 101 à 120.

Questionnaire et démarche de prévention du stress

Bien que les questionnaires et les échelles de mesure soient les méthodes les plus fréquemment utilisées pour l'évaluation des risques de stress et l'identification des conditions de travail à l'origine de situations contraignantes, l'expérience montre que les utilisateurs potentiels se posent de nombreuses questions tout au long du

TABLEAU I

Les différentes catégories de questionnaires.

Repérage par des tiers de situations stressantes
■ Évolutions et Relations en Santé au Travail (EVREST)
■ Grille d'identification des risques psychosociaux au travail – INSPQ, Québec
■ Guide d'indicateurs de dépistage des risques psychosociaux – INRS
■ Listes de contrôle de la Fondation de Dublin
■ Listes de contrôle de la SUVA PRO
Situations de travail perçues
■ Inventaire de description de l'activité professionnelle (IDAP)
■ Job Content Questionnaire (questionnaire de Karasek)
■ Job Stress Survey (JSS)
■ Nursing Stress Scale (NSS)
■ Questionnaire « Déséquilibre efforts-récompenses » (questionnaire de Siegrist)
■ Questionnaire sur le vécu du travail – Club européen de la santé
■ Questionnaire sur les facteurs de stress au travail de la Fondation de Dublin
■ Questionnaire sur les troubles musculosquelettiques de l'INRS – Facteurs psychosociaux
■ Working Conditions and Control Questionnaire (WOCCQ)
Évaluation du stress perçu (PSS)
■ Echelle de stress perçu (PSS)
■ Echelle visuelle analogique (EVA)
Stratégies d'adaptation au stress
■ Coping Inventory for Stressful Situations (CISS)
■ Ways of Coping Checklist (WCC)
Symptômes de stress
■ Item unique
■ Mesure du Stress Psychologique (MSP)
■ Questionnaire de symptômes de stress - Club Européen de la Santé
■ Questionnaire de symptômes de stress de la Fondation de Dublin
■ Questionnaire sur les troubles musculosquelettiques de l'INRS – Symptômes de stress
■ Stress Professionnel Positif et Négatif * (SPPN)
Atteinte du rapport psychologique au travail
■ Maslach Burnout Inventory (MBI)
■ Stress Professionnel Positif et Négatif * (SPPN)
Atteinte à la santé physique et mentale
■ Center for Epidemiologic Studies Depression Scale (CES-D)
■ General Health Questionnaire (GHQ)
■ Hospital Anxiety Depression Scale (HADS)
■ Mini International Neuropsychiatric Interview (MINI)
■ Nottingham Health Profile (NHP)
■ State-Trait Anxiety Inventory (STAI)
Violences internes au travail
■ Leymann Inventory of Psychological Terror (LIPT)
Questionnaires transversaux
■ Copenhagen Psychosocial Questionnaire (COPSOQ)
■ Questionnaire d'évaluation des représentations des déterminants organisationnels et psychosociaux de l'activité (QERDOPS)
■ Santé au travail, INRS et Nancy 2 (SATIN)
■ Vécu au travail (VT)

** Ce questionnaire a deux objectifs différents, c'est pourquoi il apparaît dans ces deux catégories. Cependant, il sera publié dans la catégorie « Symptômes de stress ».*

déroulement du diagnostic : quelle est la meilleure méthode pour réaliser un diagnostic ? Quels sont les intérêts et les limites d'un questionnaire ou d'une échelle de mesure ? Une fois la décision prise, quel outil choisir ? À quel moment de la démarche l'utiliser ? Comment optimiser les conditions de déroulement de l'enquête ? Comment préparer l'outil ? Y a-t-il des précautions particulières pour le traitement des données ? Comment interpréter les résultats ?… Cette troisième partie vise à aider les préventeurs, le médecin du travail et plus largement les personnes en charge de la santé au travail dans une entreprise, qui doivent décider des conditions de déroulement d'une enquête en leur proposant des premiers éléments de réponse sous la forme de conseils généraux. Le recours à une expertise complémentaire dans ce domaine peut s'avérer nécessaire.

TABLEAU II

Liste des critères utilisés dans l'analyse des outils.

Critères	Contenu des critères
Nom de l'outil	*En langue originale et sa traduction en français*
Noms des auteurs	*Noms, initiales des prénoms*
Objectifs	*Description succincte des objectifs visés par l'outil*
Année de première publication	*Année de création de la version originale*
Cadre, définition, modèle	*Définition des concepts mesurés* *Modèle théorique à la base de l'outil* *Type de population sur lequel l'outil a été validé*
Niveau d'investigation	*Pré-diagnostic ou diagnostic*
Langue d'origine	*Anglais, français, allemand…*
Traduction	*Française…*
Vocabulaire	*Difficultés particulières de compréhension*
Versions existantes	*Nombre de versions disponibles / longue ou courte / nombre d'items pour chaque version*
Structuration de l'outil	*Nombre d'échelles* *Nombre de questions par échelle*
Modalités de réponse et cotation	*Échelle de Likert* *Élaboration d'un score global ou par échelle*
Temps de passation	*Estimé par les auteurs ou les praticiens-utilisateurs*
Disponibilités et conditions d'utilisation	*Libre accès* *Autorisation des auteurs* *Licence d'utilisation (droits d'auteur)*
Qualités psychométriques	*Validité* *Fidélité* *Sensibilité*
Étalonnage	*Existence de valeurs de référence et d'un échantillon normatif*
Biais, critiques, limites	
Observations particulières	
Références bibliographiques	*Articles consultés pour documenter l'ensemble des rubriques ci-dessus*

QUELS OUTILS, À QUELS MOMENTS DE LA DÉMARCHE DE PRÉVENTION DU STRESS ?

Les outils d'évaluation du stress au travail sont à resituer dans la démarche de prévention du stress, telle qu'elle a été décrite dans la brochure INRS « *Stress au travail. Les étapes d'une démarche de prévention* ». Ces outils peuvent être utilisés à trois moments de la démarche : le pré-diagnostic, le diagnostic et l'évaluation.

Pré-diagnostic

La démarche de prévention du stress ou des risques psychosociaux commence généralement par un pré-diagnostic afin d'objectiver les problèmes, faire apparaître leur dimension collective. Cette phase a pour objectif de convaincre l'ensemble des acteurs de l'entreprise de s'engager dans la démarche.

Ce pré-diagnostic, réalisé en interne, s'appuie sur un certain nombre d'informations disponibles dans l'entreprise. Ce sont des données liées au fonctionnement de celle-ci ou à la santé de ses salariés pouvant être mobilisées assez facilement et rapidement auprès des services concernés. Leur recueil ne nécessite pas d'investigations particulières et il est facilité par l'utilisation d'outils, (catégorie « Repérage par des tiers de situations stressantes » du ***tableau I***). Le recours à un questionnaire d'auto-évaluation, renseigné par les salariés eux-mêmes, n'est pas recommandé pour cette étape de pré-diagnostic car l'objectif n'est pas immédiatement l'action mais vise à disposer d'arguments pour mettre en évidence la présence du risque. Ce pré-diagnostic peut être réalisé avec l'appui d'un groupe projet. Ce dernier sera de toute façon nécessaire à partir de la phase diagnostic.

Diagnostic approfondi

Les objectifs du diagnostic approfondi sont :

- de préciser l'ampleur du problème en évaluant le niveau de stress des salariés,
- d'identifier les déterminants liés au travail et à son organisation pouvant être à l'origine des tensions,
- et de repérer les groupes de salariés (ateliers ou services) en difficulté pour définir des cibles d'action prioritaires.



La réalisation de ce diagnostic s'appuie sur différentes méthodes complémentaires : entretiens collectifs ou individuels, observations du travail, analyse de documents administratifs, questionnaires… Les questionnaires ou échelles de mesure utilisés à cette étape sont des outils « d'auto-évaluation » pour lesquels l'avis des salariés est requis et sont énoncés dans les 8 dernières parties du tableau I.

Évaluation

Pour être complète, la démarche de prévention doit suivre et évaluer l'efficacité des actions mises en place suite au diagnostic approfondi. Cette évaluation peut être réalisée en suivant les indicateurs de dépistage des risques psychosociaux choisis et mis en place à l'étape de pré-diagnostic ou en administrant à nouveau l'outil utilisé dans la phase du diagnostic approfondi.

QUELQUES CONSEILS MÉTHODOLOGIQUES

Dans une enquête par questionnaire, si le choix de celui-ci est important, les conditions dans lesquelles il est utilisé sont primordiales. L'enquête est un processus social soumis à de nombreux biais qui, s'ils ne sont pas maîtrisés au moment de la préparation de l'enquête, compromettront la qualité des résultats, leur interprétation ou encore leur utilité sociale. Compte tenu de ces exigences, il peut être nécessaire de faire appel à des compétences extérieures qui apporteront leur expertise.

Choix de l'outil

Ce choix doit tenir compte des spécificités du travail, de ses exigences, des caractéristiques des salariés (par exemple, certains outils présentent des difficultés de vocabulaire). L'outil retenu doit pouvoir mettre en évidence les principales difficultés rencontrées par les salariés. Pour cela, son contenu doit être connu afin de retenir l'outil le mieux adapté. En cas de doute, des entretiens exploratoires permettront de cibler les caractéristiques du travail à évaluer.

Il est recommandé de ne prendre que des outils français ou traduits en français et dont les qualités psychométriques ont été vérifiées sur une population française afin d'éviter les biais culturels.

La longueur de l'outil est également un élément à prendre en compte. En effet, trop long, il peut lasser les répondants, donc nuire à la qualité des résultats ou les contraindre à répondre trop rapidement s'ils ne disposent que d'un temps restreint pour le renseigner. Enfin, il est préférable de choisir un outil pour lequel il existe un étalonnage de référence, ou sinon de recourir à une comparaison entre ateliers, entre services…

Les bonnes pratiques recommandent de retenir au moins deux outils : un premier pour évaluer le niveau de stress éprouvé par les salariés et un second pour repérer des éléments des situations de travail susceptibles d'être à l'origine de mal-être au travail ou pouvant, à terme, le devenir. Si le niveau de stress perçu est faible et que les salariés estiment se trouver dans des situations de fortes contraintes, l'intérêt est d'agir en prévention, c'est-à-dire avant que l'état de santé des salariés ne commence à se dégrader.

Une fois le (ou les) outil(s) choisi(s), une réflexion doit être menée sur la mise en place des conditions de passation (qui va être concerné ? quand se fera son administration ? qui va traiter les résultats ? …).

Rédaction des questions sociodémographiques et socio-organisationnelles

Habituellement, la première partie d'un outil est consacrée à la description des caractéristiques sociodémographiques et socio-organisationnelles des répondants : sexe, âge, atelier ou service, horaires de travail, ancienneté dans l'entreprise, au poste, type de contrat de travail… La formulation des questions doit être adaptée à l'entreprise. Cette phase de préparation nécessite de connaître le fonctionnement de l'entreprise et les principales caractéristiques de sa population. Par exemple, la variable « atelier ou service où l'on travaille » va permettre la comparaison des scores obtenus dans les différents services sur la perception des salariés de leur environnement de travail.

Information et communication au personnel

L'information des salariés tout au long du processus et la mise en place d'un plan de communication sont particulièrement importantes. Les salariés doivent être informés régulièrement, par exemple au travers des comptes rendus du CHSCT ou autre instance, dès le début des discussions entre partenaires sociaux, puis sur le déroulement des différentes étapes, à partir du moment où la décision est prise d'engager, dans l'entreprise, une démarche de prévention sur le stress au travail ou les risques psychosociaux.

Implication des salariés

La participation à l'enquête par questionnaire doit se faire sur la base du volontariat. Les modalités de retour des réponses doivent garantir l'anonymat et la confidentialité de celles-ci.

La participation des salariés est d'autant plus importante qu'ils sont convaincus de l'utilité de leurs réponses. En contrepartie de l'engagement, voire du risque qu'ils peuvent penser prendre en répondant, des garanties sur la mise en place d'un plan d'action à l'issue de l'enquête doivent leur être données. Pour indiquer l'importance accordée à une implication la plus large possible, le choix peut être fait et communiqué de fixer un taux de retour minimal en dessous duquel les réponses ne seront pas traitées, faute de représentativité.

Conditions de passation et modalités de retour des réponses

La façon de faire parvenir le questionnaire aux intéressés doit être débattue. Plusieurs voies sont possibles : par mail, remis en main propre, annexé au bulletin de paie, proposé lors de la visite médicale, administré en petits groupes sur le temps de travail…

La passation du questionnaire doit être organisée de préférence pendant le temps de travail, modalité qui garantit un meilleur taux de retour.

Si le questionnaire est renseigné manuellement, il faut prévoir un retour sous enveloppe, déposée dans une urne ou retournée à l'organisme extérieur qui se charge de l'enquête. Il est possible de mettre en place un point d'accueil pour certains participants qui souhaitent s'exprimer sur le sujet d'une autre manière, ceux qui ne comprennent pas bien les items, voire ceux qui ne savent lire qu'imparfaitement…

Traitement et interprétation des données

Le traitement des données est réalisé par une personne compétente en analyses statistiques de données quantitatives. *A minima*, deux étapes sont incontournables :

- l'étude de la représentativité des répondants par rapport à la population totale des salariés visés par l'enquête,

- la description des scores des échelles avec éventuellement comparaison aux données de référence.

Pour l'interprétation des données, il est nécessaire de prendre pleinement appui sur le groupe projet constitué pour faire le lien entre les résultats et le contexte de l'entreprise et ses conditions de travail. Sinon le diagnostic risque d'être sans suite ou de donner lieu à des actions inadaptées aux difficultés rencontrées par les salariés. La connaissance du travail reste un élément central, indispensable à l'efficacité d'une enquête par questionnaire sur le stress au travail.

Restitution des résultats

La restitution des résultats est généralement faite oralement dans un premier temps au groupe projet, puis au CHSCT, voire à l'ensemble des salariés. Compte tenu des remarques apportées lors de cette présentation, un rapport écrit est alors rédigé par l'intervenant extérieur. Le groupe projet et l'ensemble des acteurs de l'entreprise doivent pouvoir s'approprier les résultats de l'enquête pour définir ensemble les suites à donner.

Conclusion

Les questionnaires et les échelles de mesure constituent une méthode, parmi d'autres, pour évaluer le stress, les facteurs de risques et ses effets. Les méthodes qualitatives (entretiens, observations) peuvent être associées ou parfois se substituer aux questionnaires en tant que méthode de diagnostic.

Dans tous les cas, la restitution des résultats doit déboucher sur l'élaboration et la mise en œuvre d'un plan d'action qui doit être suivi et évalué. Cela passe par la mise en place d'une conduite de projet dans l'entreprise, avec la création de groupes de travail et la coordination des acteurs de prévention dans l'entreprise.

À terme, il s'agit bien d'intégrer la prévention du stress au travail dans la gestion quotidienne de la santé et de la sécurité au travail dans l'entreprise.

ÉLÉMENTS BIBLIOGRAPHIQUES, POINTS À RETENIR ET ANNEXE PAGES SUIVANTES

Éléments bibliographiques

■ BEECH JR, HARDING L - Tests, mode d'emploi. Guide de psychométrie. Traduit de l'anglais sous la direction de JP Rolland et JL Mogenet. Paris : ECPA ; 1994 : 180 p.

■ BOULETREAU A, CHOUANIÈRE D, WILD P, FONTANA JM - Concevoir, traduire et valider un questionnaire. À propos d'un exemple, EUROQUEST. Notes scientifiques et techniques de l'INRS NS 178. Paris : INRS ; 1999 : 46 p.

■ CHOUANIÈRE D, LANGEVIN V, GUIBERT A - Stress au travail. Les étapes d'une démarche de prévention. Repères pour le préventeur en entreprise. Édition INRS ED 6011. Paris : INRS ; 2007 : 31 p.

■ CHOUANIÈRE D - Dépister ou diagnostiquer les risques psychosociaux : quels outils ? *Bull Épidémiol Hebd.* 2009 ; 25-26 : 261-65.

■ DELAUNOIS M, MALCHAIRE J, PIETTE A - Classification des méthodes d'évaluation du stress en entreprise. *Méd Trav Ergon.* 2002 ;39 (1) : 13-28.

■ FRANÇOIS M, LIÉVIN D - Démarche de prévention du stress au travail. La réalisation d'un diagnostic organisationnel. Études et enquêtes TF 150. *Doc Méd Trav.* 2006 ; 107 : 309-19.

■ KLINE P – The handbook of psychological testing. 2nde edition. London : Routledge ; 2000: 744 p

■ LANGEVIN V, GUYOT S, MONTAGNEZ A, ENCONTRE D ET AL. - Prévention des risques psychosociaux. Et si vous faisiez appel à un consultant ? Édition INRS ED 6070. Paris : INRS ; 2010 : 31 p.

■ Méthodes et instruments pour une analyse ergonomique et psychosociale. Bruxelles : Service public fédéral Emploi, Travail et Concertation sociale ; 2005 : 75 p.

■ NUNNALY JC – Psychometric theory. New-York : McGraw Hill ; 1978: 701 p.

■ RODARY C, PEZET-LANGEVIN V, KALIFA C - Qualité de vie chez l'enfant : qu'est ce qu'un bon outil d'évaluation ? *Arch Pédiatr.* 2001 ; 8 (7) : 744-50.

■ VOLKOFF S (ED) - L'ergonomie et les chiffres de la santé au travail : ressources, tensions et pièges. Collection Le travail en débats, série Entreprise, travail, emploi. Collection Travail et activité humaine. Toulouse : Octarès Éditions ; 2005 : 244 p.

Points à retenir

Le questionnaire est une méthode parmi d'autres possibles permettant d'évaluer le niveau de stress en entreprise, d'identifier ses sources et de repérer ses conséquences.

Le recours à un questionnaire doit s'inscrire dans une démarche de prévention ; il ne peut l'initier à lui seul. Il doit être un des éléments d'un processus social dont chaque phase doit être discutée et préparée par les différentes parties en présence (et /ou ses représentants), compte tenu du contexte et de la culture de l'entreprise.

La connaissance de l'activité et du travail est un élément central indispensable à l'efficacité d'une enquête par questionnaire.

Dans les différents questionnaires disponibles, il est nécessaire de distinguer ceux utilisés pour le pré-diagnostic et ceux utilisés dans la phase du diagnostic. Les premiers visent à repérer d'un point de vue extérieur (en hétéro-évaluation) les situations de travail difficiles pour les salariés. Les seconds, renseignés directement par les salariés (en auto-évaluation), permettent d'identifier les sources de stress et les groupes à risque dans l'entreprise.

Généralement, les questionnaires utilisés dans la phase de diagnostic sont des échelles de mesure qui permettent d'établir des scores et de hiérarchiser les sources de stress identifiées, en vue du plan d'actions.

La construction et la validation des échelles de mesure sont un long processus qui permet d'établir leurs qualités psychométriques, validité, fidélité et sensibilité.

Trente-six questionnaires publiés en français sont analysés et vont être régulièrement publiés dans la revue *Documents pour le Médecin du Travail* sous la référence FRPS.

ANNEXE I

Qualités psychométriques des échelles de mesure

Validité

Une échelle est valide si elle mesure bien ce qu'elle est censée mesurer. Il est important de noter que, sans remettre en cause l'indispensable évaluation de la validité d'un outil, la terminologie utilisée pour nommer les différentes formes de validité varie selon les auteurs et les disciplines, ce qui ne rend pas toujours facile sa compréhension.

Validité apparente

Cette validité concerne l'acceptabilité de l'outil par les personnes cibles et résulte du jugement subjectif d'experts ou simplement des utilisateurs. Après rédaction, l'échelle doit être relue par des experts du domaine. Ils apportent une appréciation subjective sur :

- l'absence de dérive par rapport à l'objet initial,
- la compréhension des questions par la population cible,
- la non-induction des réponses,
- la pertinence des modalités de réponse,
- le choix des réponses et leur nombre,
- la non-ambiguïté du vocabulaire,
- l'utilité de toutes les questions afin de réduire au maximum la longueur de l'échelle.

D'autre part, un pré-test, pratiqué de manière rigoureuse, est indispensable pour identifier les imperfections d'une première ébauche. L'échelle doit être administrée à un échantillon représentatif de la population cible, par un enquêteur confirmé, doté de connaissances méthodologiques et scientifiques, capable de dépister les biais de réponses et de les expliquer. L'effet de halo est l'un de ces biais *(cf. encadré 1)*.

Validité de contenu

Cette validité concerne la pertinence du contenu de l'échelle. Il s'agit de déterminer si le choix des items sélectionnés pour composer l'échelle est un échantillonnage représentatif de l'ensemble des items possibles pour mesurer toutes les facettes du concept exploré. Seuls des experts de ce dernier peuvent juger de la pertinence du contenu. Le fait qu'un groupe d'experts s'accorde à dire que la validité de contenu d'une échelle est satisfaisante n'est pas suffisant. Il est nécessaire de confronter la construction de l'échelle à la structuration empirique des données.

Validité critériée ou validité empirique

Certaines échelles sont conçues pour « prédire » des aspects spécifiques des conduites humaines et de la santé des personnes. Ces aspects spécifiques sont dénommées « critères ». Dans ce cas, l'échelle est dite valide si elle prédit correctement le critère en question.

La validité critériée peut être appréciée par :

- la *validité concomitante* c'est-à-dire la corrélation entre le prédicteur et le critère qui sont mesurés en même temps (étude transversale) par le biais de plusieurs questionnaires. Des corrélations raisonnables, de l'ordre de 0,30 et au-dessus, sont considérées comme acceptables, si les effectifs des échantillons sont suffisamment importants. Par exemple, une corrélation de

ENCADRÉ 1

Effet de halo

L'effet de halo est un biais méthodologique résultant de la difficulté de la personne à répondre de façon indépendante aux différentes questions qui lui sont soumises. Cela atténue la dispersion des réponses entre les questions et peut créer artificiellement ou surestimer des corrélations entre les items.
Pour éviter cet effet de halo, deux techniques sont habituellement utilisées :
- inverser l'orientation de certains items appelés alors items inversés. Par exemple, l'échelle de stress perçu (PSS) mesure à quel point les situations de la vie ou du travail sont perçues comme menaçantes, c'est-à-dire non prévisibles, incontrôlables et pénibles. Plus le score est élevé, plus le stress perçu est important. Sept des quatorze items qui composent l'échelle sont inversés pour éviter des réponses répétitives ou l'effet de halo. Par l'exemple l'item « *Avez-vous eu le sentiment de pouvoir surmonter efficacement les changements importants qui peuvent survenir dans votre travail ?* » est un item inversé.
- « mélanger » les items composant les différentes sous-échelles d'une même échelle ou d'un inventaire afin d'éviter de créer ou renforcer artificiellement des corrélations par le simple fait de la succession des items supposés mesurer la même dimension. Cette répartition aléatoire des items est présente par exemple dans le questionnaire WOCCQ.



0,40 entre l'échelle « compensation des efforts » du questionnaire « Déséquilibre efforts - récompenses » (plus couramment appelé questionnaire de Siegrist) et la détresse psychologique mesurée par le *General Health Questionnaire* (GHQ) permet de positionner la « compensation des efforts » comme prédicteur de la détresse psychologique, sous réserve de confirmer ce lien par une étude longitudinale et par la prise en compte des variables associées ;

- la *validité prédictive*, dans laquelle le prédicteur et le critère ne sont pas mesurés au même moment (étude longitudinale). Cette méthode permet de prédire des phénomènes ultérieurs. Ce type de validité apporte beaucoup de connaissances bien qu'il faille parfois attendre un certain temps avant de pouvoir mesurer les critères. Ainsi, dans le GHQ, par exemple, la réponse positive à la question « Avez-vous pleuré plusieurs fois par jour durant les trois derniers mois ? » s'est révélée prédictive d'une tentative de suicide dans les six mois suivants.

Validité de structure ou de « construit »

De nombreuses échelles sont construites pour mesurer des concepts hypothétiques, sans existence physique, appelés « construits » (constructs) car établis pour essayer de comprendre et d'expliquer des différences interindividuelles. La validité de structure explore et confirme ou non la structure de relation entre les items.

Par exemple, le questionnaire de Karasek contient trois échelles : la demande psychologique, la latitude décisionnelle et le soutien social. Diverses procédures statistiques (analyses d'items, analyses factorielles…) permettent de s'assurer que les items de chacune de ces échelles non seulement forment bien un tout homogène mais également que chacun de ces items :

- est spécifique de cette échelle et non transversal. Que faire en effet d'une proposition ou d'un item qui permettrait de mesurer à la fois la demande psychologique, la latitude décisionnelle et le soutien social ?

- contribue suffisamment à la mesure de la dimension ciblée par l'échelle (corrélation attendue supérieure ou égale à 0,40 entre chaque item et le score moyen observé sur l'échelle).

Validité de structure interne

La validité de structure interne doit confirmer, par les données, la structure postulée de l'échelle. Le lien existant entre deux items appartenant à un même domaine, donc explorant sur le plan théorique le même concept, doit être plus fort que le lien pouvant exister entre deux items de domaines différents. Cette validité s'explore classiquement par :

- les analyses factorielles utilisées pour identifier les dimensions sur lesquelles peuvent se caractériser et se distinguer les individus. Il s'agit de réduire l'information sans trop en perdre. Les procédures mathématiques sont complexes et nécessitent le recours à des logiciels statistiques. Le principe repose sur la notion de « structure latente », c'est-à-dire la présence d'un certain nombre de facteurs (ou de dimensions sous-jacentes) permettant d'expliquer pourquoi certaines des variables sont intercorrélées, alors que d'autres variables ne le sont pas *(encadré 2)*. La liaison entre un facteur et un item ou une dimension s'exprime par un coefficient de saturation ou de contribution ;

- le modèle de Rasch est un cas particulier de modèle de réponse à l'item. Il n'est approprié que pour la construction et l'analyse de variables unidimensionnelles où tous les items sont censés mesurer le même « construit ». Le modèle de Rasch postule que les items ont un même pouvoir discriminatif mais un niveau de difficulté distinct. Dans ce cadre, contrairement à la théorie classique des tests, la théorie de réponse à l'item ne se soucie pas des scores obtenus sur des échantillons aléatoires mais se base sur les réponses individuelles à des items particuliers. Chaque item et chaque réponse du sujet à l'item sont donc considérés séparément comme une source d'informations à propos de l'échelle. Le modèle de Rasch s'inspire des travaux sur les tests d'aptitude de Guttman, ce dernier ayant mis au point une échelle permettant, sous certaines conditions strictes, de classer les sujets par compétence croissante sur un continuum et par ordre de difficultés des items.

Validité de structure externe

La validité de structure externe, également appelée analyse multitraits multiméthodes, nécessite de disposer de deux concepts différents et de deux méthodes distinctes pour mesurer chacun de ces concepts (par exemple un entretien structuré et une échelle). L'examen de la matrice de corrélations entre les quatre séries de scores obtenus permet d'estimer à la fois la validité convergente et divergente des deux concepts :

- la validité convergente est testée par la corrélation entre les mesures du même concept par les deux méthodes différentes. Cette corrélation doit être positive et très élevée ;

- la validité divergente est plus exigeante. La corrélation entre les mesures du même concept estimé par les deux méthodes différentes doit non seulement être très élevée, mais également être nettement supérieure à la corrélation entre les deux concepts différents mesurés par la même méthode. Cette approche permet de démontrer que les scores obtenus sont avant tout fonction du concept et non pas de la méthode d'évaluation.

ENCADRÉ 2

Les analyses factorielles

Il existe deux types d'analyses factorielles en fonction de la connaissance *a priori* du nombre de facteurs sous-jacents.
L'***analyse factorielle exploratoire*** est employée quand le nombre de facteurs sous-jacents n'est *a priori* pas connu ; elle permet de les identifier et procède en deux temps distincts :
- l'extraction permet de déterminer le nombre de facteurs pertinents à considérer ;
- la rotation est une redéfinition de la composition des facteurs afin d'en faciliter l'interprétation (rotation « varimax », qui est une méthode automatique dont le principe directeur est de maximiser la variance des saturations de chaque facteur).

Les analyses factorielles avec rotation « varimax » dégagent des axes orthogonaux. C'est une méthode automatique, dont le principe directeur est de maximiser la variance des corrélations avec chaque facteur. Elle est adaptée lorsque les dimensions sous-jacentes sont indépendantes ou lorsqu'il n'y a pas (ou que cela n'est pas souhaitable) d'hypothèse *a priori* sur le regroupement des variables.

L'***analyse factorielle confirmatoire*** est utilisée pour conforter (ou non) les hypothèses formulées sur le nombre de facteurs sous-jacents existants au travers des données. Elle permet donc de tester un modèle théorique explicite formulé *a priori* par le chercheur. Celui-ci doit identifier à l'avance les facteurs faisant partie du modèle théorique qu'il veut mettre à l'épreuve et sélectionner les variables observées contribuant à chaque facteur. Selon la règle de Nunnally (1978) pour confirmer les hypothèses posées, il faut que :
- chaque item contribue à la dimension attendue, avec une valeur supérieure ou égale à 0,40,
- la contribution sur la dimension attendue dépasse de 0,20 les contributions de l'item sur les autres dimensions.

Validité concourante

Il n'existe pas en psychologie d'outil absolument parfait et infaillible. Cependant, certaines échelles se sont imposées comme des outils de référence (« gold standard ») au fil des multiples études de validité dont elles ont fait l'objet.

Ainsi, une manière de s'assurer qu'une nouvelle échelle mesure bien ce qu'elle est sensée mesurer consiste à vérifier la corrélation qu'elle entretient avec un outil de référence mesurant la même caractéristique et présentant des qualités psychométriques bien établies.

Validité discriminante

La validité discriminante porte sur la capacité d'une échelle à pouvoir distinguer des groupes à partir de la caractéristique étudiée quand on suppose qu'ils se différencient effectivement sur celle-ci. La validité discriminante équivaut ici quasiment à la spécificité d'une échelle de mesure (voir la partie sur la sensibilité).

Fidélité ou fiabilité

Une échelle est fidèle ou fiable si les informations qu'elle produit sont indépendantes de celui qui effectue la mesure, du moment de la mesure, du choix de l'échantillonnage des questions. Plus un outil de mesure est précis dans la mesure de ce qu'il est censé produire, plus les informations recueillies peuvent être considérées comme fiables.

Données indépendantes de l'enquêteur

Fidélité interjuge

Elle vérifie que les réponses de l'enquêté sont indépendantes de l'enquêteur. Elle s'étudie en calculant des coefficients de concordance Kappa ou des coefficients de corrélation intraclasses (CCI) obtenues entre deux ou plusieurs enquêteurs avec le même instrument.

Fidélité intrajuge

Elle vérifie que le codage d'une même séquence réalisé par l'enquêteur ne varie pas dans le temps. Elle est surtout utilisée dans le codage d'observations et s'appelle aussi constance (ou stabilité) des codages. Comme précédemment, les coefficients Kappa ou les CCI peuvent être calculés.

Données indépendantes du moment de l'enquête : la méthode test-retest

Selon la stabilité des dimensions mesurées, les résultats à une échelle doivent être plus ou moins constants ou stables. Cette constance est évaluée au moyen de la méthode test-retest qui consiste à soumettre deux fois la même échelle au même échantillon de sujets après un intervalle de temps plus ou moins long. Une corrélation appelée coefficient de fidélité est calculée entre les résultats des deux passations. Plus ce coefficient de fidélité est élevé, plus la mesure est stable. Au contraire, plus le coefficient de fidélité est faible, moins la mesure est stable et elle ne permet donc pas de caractériser les sujets de façon fiable.

Pour des dimensions relativement stables, la fidélité test-retest à court terme (deux à trois semaines) suppose un coefficient de fidélité, de 0,80 à 0,90.

À long terme (deux mois ou plus), un coefficient de 0,60 est satisfaisant et suggère que l'échelle est effectivement stable dans le temps.

Données indépendantes de l'instrument de mesure

Méthode des formes parallèles

Cette méthode est utilisée quand il faut évaluer à plusieurs reprises les mêmes personnes, par exemple à différentes phases de traitement, et que des effets d'apprentissage, de mémoire ou de monotonie risquent d'influencer les réponses au retest.

Pour vérifier l'équivalence, les deux formes de l'échelle sont administrées aux mêmes sujets d'un échantillon et un coefficient de corrélation ou « coefficient d'équivalence » entre les deux est calculé. Un coefficient élevé signifie que le classement des sujets est le même pour les deux formes du test, et donc que les variations dues aux questions ou items sont négligeables. Un coefficient faible indique que le classement des sujets varie selon les items. Dans ce cas, les deux formes ne sont pas équivalentes, l'interprétation des scores devient ambiguë et aucune des deux formes ne peut être retenue.

Par exemple, l'outil MSP (mesure du stress psychologique) dispose de deux formes parallèles.

Consistance ou cohérence interne ou homogénéité

Il s'agit de vérifier que les mesures répétées à l'intérieur d'une même échelle (plusieurs items pour mesurer une même dimension) sont convergentes, c'est-à-dire qu'elles vont dans le même sens. On parle alors de vérification de la consistance ou homogénéité interne de l'instrument de mesure : les items de chaque dimension doivent former un « tout cohérent ».

Cette consistance interne se mesure de plusieurs façons :

- Le coefficient de consistance interne ***alpha de Cronbach*** correspond approximativement à la corrélation moyenne entre les items, pondérée par le nombre d'items. Une valeur supérieure à 0,70 correspond à un échantillonnage homogène des items. Cet indicateur statistique présente néanmoins l'inconvénient d'être sensible à la longueur de l'échelle : de par son mode de calcul, ce coefficient peut avoir tendance à augmenter artificiellement quand le nombre d'items est élevé.

- La ***méthode « split-half »*** consiste, après la passation de l'échelle à un groupe de personnes, à répartir les items en deux moitiés équivalentes. Ensuite, il s'agit de mesurer la corrélation entre les deux moitiés.

- L'***analyse d'items*** peut servir de base pour restructurer l'échelle, la repenser globalement afin qu'elle présente certaines caractéristiques souhaitées. La corrélation inter-items et la corrélation item-tout sont les indices principaux de l'analyse d'items. En effet, même s'il s'agit d'une échelle unidimensionnelle, il peut arriver que des items aient entre eux des corrélations plus fortes, et moins fortes avec d'autres items, reflétant ainsi différentes facettes de la caractéristique évaluée.

Sensibilité

La sensibilité d'une échelle de mesure est sa capacité à pouvoir faire apparaître des différences entre les individus (ou les groupes) ou des changements dans le temps (dans le cas de l'étude d'une caractéristique ou d'un phénomène évolutif). La mesure doit donc couvrir tout le champ des valeurs possibles pour les personnes, appelé zone de performance.



Sensibilité au changement

L'échelle de mesure doit être apte à mesurer les changements, afin de mettre en évidence, par exemple, une amélioration des symptômes de stress suite à des actions de prévention en entreprise. Ceci suppose qu'une évaluation par la même échelle de mesure soit réalisée avant (pré) et après (post) la mise en place de l'action, afin de mesurer le changement.

S'il est vrai que l'utilisation des indices pour calculer la fidélité d'une échelle est largement consensuelle, il n'en est pas de même pour les indices de sensibilité au changement. Cependant, les indicateurs basés sur la distribution statistique des observations, comme la réponse moyenne standardisée (*Standardized Response Means*) ou la taille de l'effet (*effect size*), permettent de quantifier l'importance du changement observé.

Il existe d'autres indices, basés sur des mesures de changement spontané recueillies rétrospectivement (incluant une mesure d'impression globale de changement).

Sensibilité aux différences interindividuelles ou intergroupes

Ici, une échelle de mesure est sensible si elle mesure le phénomène étudié avec une finesse suffisante pour permettre de distinguer les individus ou groupes d'individus. Les statistiques descriptives telles que l'histogramme des réponses, les effets plancher et plafond ou les indicateurs de tendance centrale et de dispersion (moyenne, étendue, écart-type) permettent d'apprécier la sensibilité d'une échelle de mesure.

L'exemple d'un thermomètre pour la température corporelle permet de bien comprendre cette qualité psychométrique :

- si le thermomètre ne distingue que les demi-degrés (36°, 36,5°, 37°, 37,5°…), la mesure est trop grossière et ne distingue pas de petites différences entre sujets ou pour un même sujet au cours du temps. Ce thermomètre manque de précision ;

- si le thermomètre ne couvre que la zone des valeurs inférieures ou égales à 38°, il ne permet pas de distinguer les personnes ayant une température supérieure à 38° qui sont toutes au plafond de la zone des valeurs étudiées. Il s'agit d'un effet plafond ;

- si le thermomètre ne couvre que la zone des valeurs supérieures ou égales à 38°, les personnes ayant une température inférieure à 38° ne sont pas distinguées de celles dont la température est de 38°. Il s'agit d'un effet « plancher ».

Il arrive parfois que le terme sensibilité soit utilisé dans un sens différent de celui mentionné ci-dessus. En effet, en épidémiologie, on peut étudier les propriétés métriques d'un questionnaire servant au dépistage individuel. On fait alors référence à une sensibilité qui renvoie à la capacité de l'échelle à reconnaître comme telles les personnes présentant une pathologie, un trouble, une caractéristique personnelle…(vrais positifs), tandis que la spécificité renvoie à la capacité de l'échelle de mesure à reconnaître comme telles les personnes ne présentant pas cette pathologie, ce trouble, cette caractéristique personnelle…(vrais négatifs).